



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

WHS

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/649,070	08/27/2003	Ara Victor Nefian	ITL.1033US (P16814)	4642
21906 7590 09/26/2007 TROP PRUNER & HU, PC 1616 S. VOSS ROAD, SUITE 750 HOUSTON, TX 77057-2631			EXAMINER CARTER, AARON W	
			ART UNIT 2624	PAPER NUMBER
			MAIL DATE .09/26/2007	DELIVERY MODE PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/649,070

Applicant(s)

NEFIAN ET AL.

Examiner

Aaron W. Carter

Art Unit

2624

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 13 July 2007.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 7, 8, 10-17, 20, 21, 24 and 25 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 7, 8, 10-17, 20, 21, 24 and 25 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 27 August 2003 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This action is responsive to papers filed on July 13, 2007.

Response to Amendment

2. In response to applicant's amendment received on July 13, 2007, all requested changes to the claims have been entered. Claims 24 and 25 have been added. Claims 1-6, 9, 18-19 and 22-23 have been cancelled.

Response to Arguments

3. Applicant's arguments with respect to claims 7, 16 and 20 have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 7, 8, 10-13 and 15 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu, already of record, in view of the article Embedded Bayesian Networks for Face Recognition to Nefian, already of record.

As to claim 7, Basu discloses a method comprising:

Recognizing a face of a subject from first entries in a database (*column 6, lines 32-58, wherein the facial recognition of an individual speaking is performed*);

Recognizing audio-visual speech of the subject from second entries in the database (*column 5, lines 28-42 and column 11, lines 10-31, wherein speech is recognized using the audio of a audio-visual signal and that speech is verified using the utterance verifier, which uses the video of the audio-visual signal*); and

Identifying the subject based on recognizing the face and recognizing the audio-visual speech (*column 8, lines 43-50, wherein the recognition results determined by the face and audio-visual speech recognizers are combined to produce a joint score which is used to identify the individual*).

Basu does not disclose expressly including modeling an image including the face using an embedded hidden Markov model (EHMM), wherein the EHMM is a hierarchical statistical model having a parent layer corresponding to a super state of the EHMM and including a plurality of nodes to represent hidden nodes, each node referring to a plurality of second nodes of a child layer corresponding to a state of the EHMM, the plurality of second nodes each referring to an observation node, and wherein the state of the EHMM is described by a mixture of a plurality of Gaussian density functions having diagonal covariance matrices.

However, Nefian discloses recognizing a face of a subject from first entries in a database (Abstract), including modeling an image including the face using an embedded hidden Markov model (EHMM), wherein the EHMM is a hierarchical statistical model having a parent layer

Art Unit: 2624

corresponding to a super state of the EHMM and including a plurality of nodes to represent hidden nodes, each node referring to a plurality of second nodes of a child layer corresponding to a state of the EHMM, the plurality of second nodes each referring to an observation node and wherein the state of the EHMM is described by a mixture of a plurality of Gaussian density functions having diagonal covariance matrices (*Fig. 2a, equation 1 and Section 3, wherein square nodes correspond to the parent node and the circle nodes correspond to the child nodes or observation nodes and equation 1 corresponds to the mixture of Gaussian density functions having diagonal covariance matrices*).

Basu & Nefian are combinable because they are from the same art of image processing, specifically recognition.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to use the process of modeling faces for face recognition, as taught by Nefian, with the process of face recognition and audio-visual speech recognition for identifying a subject, as disclosed by Basu.

The suggestion/motivation for doing so would have been to provide an increased recognition rate (Nefian, Table 2).

Therefore, it would have been obvious to combine Basu with Nefian to obtain the invention as specified in claim 7.

As to claim 8, the combination of Basu and Nefian discloses the method of claim 7, further comprising providing the subject access to a restricted area after identifying the subject (*Basu, column 11, lines 60-65*).

As to claim 10, the combination of Basu and Nefian discloses the method of claim 7, further comprising obtaining observation vectors from a sampling window of the image (*Nefian, section 4*).

As to claim 11, the combination of Basu and Nefian disclose the method of claim 10, wherein the observation vectors comprise discrete cosine transform coefficients (*Nefian, section 4*).

As to claim 12, the combination of Basu and Nefian disclose the method claim 7, wherein recognizing the face comprises performing a Viterbi decoding algorithm (*Nefian, section 5*).

As to claim 13, the combination of Basu and Nefian discloses the method of claim 7, wherein recognizing the audio-visual speech further comprises detecting and tracking a mouth region using vector machine classifiers (*Basu, column 10, lines 14-25*).

As to claim 15, the combination of Basu and Nefian discloses the method of claim 7, further comprising results of recognizing the face and recognizing the audio-visual speech pattern according to a predetermined weighting to identify the subject (*Basu, column 8, lines 43-50*).

Art Unit: 2624

6. Claims 16, 17, 20 and 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu in view of the article "A Bayesian Approach to Audio-Visual Speaker Identification" to Nefian et al. ("Nefian2") (already of record).

As to claim 16, Basu discloses a system comprising:

At least one capture device to capture audio-visual information from a subject (*column 4, lines 15-23 and column 13, lines 31-67*);

A first storage device coupled to the at least one capture device to store code to enable the system (*column 13, lines 31-67*) to recognize a face of the subject from first entries in a database (*column 6, lines 32-58, wherein the facial recognition of an individual speaking is performed*), recognize audio-visual speech of the subject from second entries in the database (*column 5, lines 28-42 and column 11, lines 10-31, wherein speech is recognized using the audio of a audio-visual signal and that speech is verified using the utterance verifier, which uses the video of the audio-visual signal*), and identify the subject based on the face and the audio-speech (*column 8, lines 43-50, wherein the recognition results determined by the face and audio-visual speech recognizers are combined to produce a joint score which is used to identify the individual*); and

A processor coupled to the first storage to execute the code (*column 13, lines 31-67*).

Basu does not disclose expressly, model an image including the face using an embedded hidden Markov model, model the image and an audio sample using a coupled hidden Markov model, and identify the subject based on the face and the audio-speech according to $\lambda_f L(O^f|k) + \lambda_{av} L(O^a, O^v|k)$, where O^a , O^v and O^f are audio speech, visual speech and face of the captured

audio-visual information, $L(*|k)$ is an observation likelihood for a k^{th} entry in the database, and $\lambda_f \lambda_{av} > 0$, $\lambda_f + \lambda_{av} = 1$ are weighting coefficients for face and audio-visual speech likelihood of recognition.

However, Nefian2 discloses to recognize a face of the subject from first entries in a database, model an image including the face using an embedded hidden Markov model (*Section 2*), model the image and an audio sample using a coupled hidden Markov model (*Section 3*), recognize audio-visual speech of the subject from second entries in the database (*Section 3*), and identify the subject based on the face and the audio-speech according to $\lambda_f L(O^f|k) + \lambda_{av} L(O^a, O^v|k)$, where O^a , O^v and O^f are audio speech, visual speech and face of the captured audio-visual information, $L(*|k)$ is an observation likelihood for a k^{th} entry in the database, and $\lambda_f \lambda_{av} > 0$, $\lambda_f + \lambda_{av} = 1$ are weighting coefficients for face and audio-visual speech likelihood of recognition (*Section 5*); and

Basu & Nefian2 are combinable because they are from the same art of image processing, specifically recognition.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to use the process of modeling the face using an EHMM and the image and audio sample using a CHMM, as well as, using the identification process, as taught by Nefian2, with the system disclosed by Basu.

The suggestion/motivation for doing so would have been to provide improved identification accuracy (Nefian2, Abstract).

Therefore, it would have been obvious to combine Basu with Nefian2 to obtain the invention as specified in claim 16.

Art Unit: 2624

As to claim 17, the combination of Basu and Nefian2 discloses the system of claim 16, wherein the database is stored in the first storage device (*Basu, column 13, lines 31-67 and column 3, lines 45-48*).

As to claim 20, please refer to the rejection of claim 16 above.

As to claim 21, the article of claim 20, further comprising instructions that if executed enable the system to provide the subject access to a restricted area after the subject is identified (*Basu, column 11, lines 60-65*).

7. Claim 14 is rejected under 35 U.S.C. 103(a) as being unpatentable over Basu and Nefian as applied to claim 7 above, and further in view of Nefian2.

As to claim 14, the combination of Basu and Nefian discloses the method of claim 7.

The combination of Basu and Nefian does not disclose expressly wherein recognizing the audio-visual speech comprises modeling an image and an audio sample using a coupled hidden Markov model.

Nefian2 discloses to recognize a face of the subject from first entries in a database, model an image including the face using an embedded hidden Markov model (*Section 2*), model the image and an audio sample using a coupled hidden Markov model (*Section 3*), recognize audio-visual speech of the subject from second entries in the database (*Section 3*), and identify the subject based on the face and the audio-speech (*Section 5*).

Art Unit: 2624

Basu, Nefian & Nefian2 are combinable because they are from the same art of image processing, specifically recognition.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to model an image and an audio sample using a coupled hidden Markov model, as taught by Nefian2, in the audio-visual speech recognition process disclosed by the combination of Basu and Nefian.

The suggestion/motivation for doing so would have been to provide improved recognition accuracy (*Nefian2, Abstract*).

Therefore, it would have been obvious to combine Basu and Nefian with Nefian2 to obtain the invention as specified in claim 14.

8. Claim 24 is rejected under 35 U.S.C. 103(a) as being unpatentable over Basu and Nefian as applied to claim 15 above, and further in view of Nefian2.

As to claim 24, the combination of Basu and Nefian discloses the method of claim 15.

The combination of Basu and Nefian does not disclose expressly wherein the predetermined weighting corresponds to $\lambda_f L(O^f|k) + \lambda_{av} L(O^a, O^v|k)$, where O^a , O^v and O^f are audio speech, visual speech and face of the captured audio-visual information, $L(*|k)$ is an observation likelihood for a k^{th} entry in the database, and $\lambda_f \lambda_{av} \geq 0$, $\lambda_f + \lambda_{av} = 1$ are weighting coefficients for face and audio-visual speech likelihood of recognition.

However, Nefian2 discloses to recognize a face of the subject from first entries in a database, model an image including the face using an embedded hidden Markov model (*Section*

Art Unit: 2624

2), model the image and an audio sample using a coupled hidden Markov model (*Section 3*), recognize audio-visual speech of the subject from second entries in the database (*Section 3*), and identify the subject based on the face and the audio-speech according to $\lambda_f L(O^f|k) + \lambda_{av} L(O^a, O^v|k)$, where O^a , O^v and O^f are audio speech, visual speech and face of the captured audio-visual information, $L(*|k)$ is an observation likelihood for a k^{th} entry in the database, and $\lambda_f \lambda_{av} \geq 0$, $\lambda_f + \lambda_{av} = 1$ are weighting coefficients for face and audio-visual speech likelihood of recognition (*Section 5*); and

Basu, Nefian & Nefian2 are combinable because they are from the same art of image processing, specifically recognition.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to include the process of modeling the face using an EHMM and the image and audio sample using a CHMM, as well as, using the identification process, as taught by Nefian2, with the system disclosed by the combination of Basu and Nefian.

The suggestion/motivation for doing so would have been to provide improved recognition accuracy (*Nefian2, Abstract*).

Therefore, it would have been obvious to combine Basu and Nefian with Nefian2 to obtain the invention as specified in claim 24.

9. Claim 25 is rejected under 35 U.S.C. 103(a) as being unpatentable over the combination of Basu and Nefian2 as applied to claim 20 above, and further in view of Nefian.

As to claim 25, the combination of Basu and Nefian2 discloses the article of claim 20, wherein the embedded hidden Markov model (EHMM) is a hierarchical statistical model having

Art Unit: 2624

a parent layer corresponding to a super state of the EHMM and including a plurality of nodes to represent hidden nodes, each node referring to a plurality of second nodes of a child layer corresponding to a state of the EHMM, the plurality of second nodes each referring to an observation node (*Nefian2, section 2*).

The combination of Basu and Nefian2 does not disclose expressly wherein the state of the EHMM is described by a mixture of a plurality of Gaussian density functions having diagonal covariance matrices.

However, Nefian discloses recognizing a face of a subject from first entries in a database (Abstract), including modeling an image including the face using an embedded hidden Markov model (EHMM), wherein the EHMM is a hierarchical statistical model having a parent layer corresponding to a super state of the EHMM and including a plurality of nodes to represent hidden nodes, each node referring to a plurality of second nodes of a child layer corresponding to a state of the EHMM, the plurality of second nodes each referring to an observation node and wherein the state of the EHMM is described by a mixture of a plurality of Gaussian density functions having diagonal covariance matrices (*Fig. 2a, equation 1 and Section 3, wherein square nodes correspond to the parent node and the circle nodes correspond to the child nodes or observation nodes and equation 1 corresponds to the mixture of Gaussian density functions having diagonal covariance matrices*).

Basu, Nefian2 & Nefian are combinable because they are from the same art of image processing, specifically recognition.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to use the process of modeling faces for face recognition, as taught by Nefian, with the

Art Unit: 2624

process of face recognition and audio-visual speech recognition for identifying a subject, as disclosed by the combination of Basu and Nefian2.

The suggestion/motivation for doing so would have been to provide an increased recognition rate (Nefian, Table 2).

Therefore, it would have been obvious to combine Basu and Nefian2 with Nefian to obtain the invention as specified in claim 25.

Conclusion

10. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the date of this final action.


Art Unit: 2624

11. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Aaron W. Carter whose telephone number is (571) 272-7445.

The examiner can normally be reached on 8am - 4:30 am (Mon. - Fri.).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Bhavesh Mehta can be reached on (571) 272-7453. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.



Aaron Carter
AU 2624